Curvature of designed armadillo repeat proteins allows modular peptide binding

Hansen, Simon ; Ernst, Patrick ; König, Sebastian L B ; Reichen, Christian ; Ewald, Christina ; Nettels, Daniel ; Mittl, Peer R E ; Schuler, Benjamin ; Plückthun, Andreas

Abstract: Designed armadillo repeat proteins (dArmRPs) were developed to create a modular peptide binding technology where each of the structural repeats binds two residues of the target peptide. An essential prerequisite for such a technology is a dArmRP geometry that matches the peptide bond length. To this end, we determined a large set (n=27) of dArmRP X-ray structures, of which 12 were previously unpublished, to calculate curvature parameters that define their geometry. Our analysis shows that consensus dArmRPs exhibit curvatures close to the optimal range for modular peptide recognition. Binding of peptide ligands can induce a curvature within the desired range, as confirmed by single-molecule FRET experiments in solution. On the other hand, computationally designed ArmRPs, where side chains have been chosen with the intention to optimally fit into a geometrically optimized backbone, turned out to be more divergent in reality, and thus not suitable for continuous peptide binding. Furthermore, we show that the formation of a crystal lattice can induce small but significant deviations from the curvature adopted in solution, which can interfere with the evaluation of repeat protein scaffolds when high accuracy is required. This study corroborates the suitability of consensus dArmRPs as a scaffold for the development of modular peptide binders.

DOI: https://doi.org/10.1016/j.jsb.2017.08.009

Curvature of designed armadillo repeat proteins allows modular peptide binding

Authors: Simon Hansen[a,1 †], Patrick Ernst[a, †], Sebastian L.B. König[a], Christian Reichen[a,2], Christina Ewald[a,3], Daniel Nettels[a], Peer R.E. Mittl[a], Benjamin Schuler[a], Andreas Plückthun[a, *]

Affiliation: [a] Department of Biochemistry, University Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

[†] equal contribution

[1] Present address: Department of Early Discovery Biochemistry, Genentech, 1 DNA Way, South San Francisco, CA 94080, USA

[2] Present address: Molecular Partners AG, Wagistrasse 14, 8952 Zürich-Schlieren, Switzerland

[3] Present address: Flow Cytometry Facility, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

Correspondence: *Andreas Plückthun Tel. +41-44-635 5570, Fax. +41-44-635 5712 e-mail: plueckthun@bioc.uzh.ch

Manuscript: Total pages: 28

Supporting information:

- Supplementary figures and tables: Fig. S1-S4; Tables ST1 & ST2 (12 pages)
- Supplementary raw data table: Curvature parameters (Excel file)

Tables: 1

Figures: 6

Coordinates: Coordinates and structure factors have been deposited at the PDB with accession codes 5MFB, 5MFE, 5MFF, 5MFG, 5MFH, 5MFI, 5MFJ, 5MFK, 5MFL, 5MFM, 5MFN and 5MFO

Abstract

Designed armadillo repeat proteins (dArmRP) were developed to create a modular peptide binding technology where each of the structural repeats binds two residues of the target peptide. An essential prerequisite for such a technology is a dArmRP geometry that matches the peptide bond length. To this end, we determined a large set (n=27) of dArmRP X-ray structures, of which 14 were previously unpublished, to calculate curvature parameters that define their geometry. Our analysis shows that consensus dArmRPs exhibit curvatures close to the optimal range for modular peptide recognition. Binding of peptide ligands can induce a curvature within the desired range, as confirmed by single-molecule FRET experiments in solution. On the other hand, computationally designed ArmRPs, where side chains have been chosen with the intention to optimally fit into a geometrically optimized backbone, turned out to be more divergent in reality, and thus not suitable for continuous peptide binding. Furthermore, we show that the formation of a crystal lattice can induce small but significant deviations from the curvature adopted in solution, which can interfere with the evaluation of repeat protein scaffolds when high accuracy is required. This study corroborates the suitability of consensus dArmRP as a scaffold for the development of modular peptide binders.

Keywords

Armadillo repeat proteins, peptide binding, protein crystallography, protein engineering, single-molecule FRET

Abbreviations

AU: asymmetric unit, CoM: center of mass, dArmRP: designed armadillo repeat proteins, nArmRP: natural armadillo repeat proteins

Introduction

Armadillo repeat proteins consist of a repetitive protein fold found in many proteins, most prominently in importin-α and β-catenin, which are involved in cellular processes including nuclear import, cell adhesion and signaling (Tewari et al., 2010). The name derives from the appearance of a *Drosophila* mutant with a defect in a segment polarity gene later identified as β-catenin (Nüsslein-Volhard and Wieschaus, 1980). Natural armadillo repeat proteins (nArmRP) are abundant in the human genome and usually consist of 4-12 consecutive repeats (Gul et al., 2016) where the repeats at either terminus (called N- and C-cap, respectively) are modified to protect the hydrophobic core from solvent exposure. Each repeat is usually made from 42 residues that fold into three triangularly arranged helices (H1, H2 and H3), and the angular stacking of several repeats creates a binding groove of adjacent H3 helices (Huber et al., 1997). The best described members of this protein family, importin-α and β-catenin, were shown to bind peptides in a stretched conformation in their binding groove (Conti et al., 1998; Graham et al., 2000). This stretched conformation of the bound peptide is enforced by bidentate hydrogen bonds between the backbone NH and O atoms of every second peptide bond and conserved asparagine residues on H3 (N^{37}, superscripted numbers refer to residue numbering within one repeat), while the specificity is established by interactions with the peptide side chains.

nArmRPs bind peptide ligands of up to six amino acid units in length in a modular fashion, i.e., two residues per repeat are recognized (Conti and Kuriyan, 2000; Conti et al., 1998). Longer modular peptide binding has not been observed in nArmRP, because they do not possess a regular curvature that fits the binding register dictated by the peptide bond lengths (Reichen et al., 2016b). Designed armadillo repeat proteins (dArmRP) were developed to expand this modular peptide recognition by engineering a completely regular scaffold that fits

the required register along the whole scaffold length and would therefore allow binding of, in theory, infinitely long peptides (Alfarano et al., 2012; Madhurantakam et al., 2012; Parmeggiani et al., 2008) (Fig. 1a). Based on this scaffold, surface-randomized repeats might be developed that specifically bind a certain dipeptide in the context of a longer peptide, and most likely this can be achieved by directed evolution methods. Such preselected repeats could then in turn be assembled to create specific peptide binders against novel primary peptide sequences without additional selections (Reichen et al., 2014a).

The correct curvature of the scaffold is a fundamental prerequisite for the development of such a technology. The curvature of a repeat protein can be described by three parameters that define a superhelix which is formed by solenoid proteins: The rise (h), the angle (2Ω) between neighboring repeats and the radius of the superhelix (r). For the parametrization of experimental structures, these values are usually calculated between the center of mass (CoM) of each repeat and the central axis of the superhelix (DiMaio et al., 2011; Kobe and Kajava, 2000; Park et al., 2015). A peptide in an extended conformation has a distance of 6.7-7.0 Å between the C^{α} atom of one residue and the C^{α} two residues further toward the C-terminus (termed C^{α}(P/P+2) distance). This distance was predicted by modeling of peptides in a relaxed β-strand conformations and was also found in peptides bound by nArmRP and dArmRP (Reichen et al., 2016b), albeit only for short stretches.

dArmRPs and bound peptides can be described as two intertwined helices with different radii (r_A and r_P, respectively) but with identical angle and rise (Fig. 1b). Therefore, the analysis of dArmRP superhelix curvature parameters permits one to assess whether or not they match the register given by the target peptide and whether they would allow for continuous modular binding.

Previously, a computational design approach was described, where predefined curvature restraints were used to design this protein scaffold. These restraints were obtained from a

model based on two repeats of yeast importin-α which exhibited a C^{α}(P/P+2) distance within the optimal range for continuous peptide binding. Still, the obtained scaffold deviated from the intended curvature, and hence, solely allowed limited modular binding over approximately three consecutive repeats (Reichen et al., 2016b). On the other hand, a consensus design approach resulted in high-affinity binding to peptides with repetitive arginine-lysine sequence ($(KR)_n$) that showed in-solution binding behavior that is consistent with modular binding (Hansen et al., 2016). However, these results still do not prove an optimal curvature of the scaffold since the peptides used were at most ten residues in length. For even longer peptides small deviations of the scaffold might add up and at some point prevent modular peptide binding.

Here we analyze 27 experimental X-ray structures of consensus and computational dArmRPs with a focus on their curvature parameters. This analysis allows us to draw conclusions regarding which design approach provides a suitable scaffold for modular peptide binding. The analysis also revealed that due to the intrinsic flexibility of dArmRP, packing into a crystal lattice can introduce geometric differences between structures of dArmRP that also influences their peptide binding mode. This hampers the characterization of designed scaffolds by crystallography alone, which was therefore supplemented by in-solution FRET measurements.

Material and methods

Cloning

Cloning of the dArmRP genes has been described previously (Madhurantakam et al., 2012; Reichen et al., 2016b). Direct fusions of (KR)-peptides to the dArmRP were introduced by PCR using oligonucleotides encoding the (KR)-peptide, a GS-linker and an overlap with

either the N- or C-terminus of dArmRP. All sequences of the different types of dArmRP used in this study are shown in Fig. 2. Constructs were subcloned with BamHI and HindIII restriction enzymes (FastDigest enzymes, Fermentas) into the vector pQE30LIC_3C, which contains a 3C-protease-cleavable MRGSHis$_6$-tag (Hansen et al., 2016). All construct sequences were confirmed by DNA sequencing (Microsynth), and glycerol stocks of the respective plasmids were prepared (overnight culture of transformed *E. coli* XL1 Blue or BL21 (DE3) with 15% glycerol, stored at -80°C).

Protein expression and purification

Protein expression was carried out in 1 l of 2xYT medium (containing 100 µg/l ampicillin and 0.5% glucose). Overnight cultures (25 ml) were inoculated from glycerol stocks, grown at 37°C and used for inoculation. The expression culture was grown to an OD_{600} of 0.7, followed by induction with 750 µM of isopropyl-β-D-thiogalactopyranoside (IPTG) and incubation at 37°C for 5 h. Cells were centrifuged (5,000 *g*, 5 min), resuspended in 25 ml TBS_W (50 mM Tris pH 8.0, 400 mM NaCl, 20 mM imidazole) and flash-frozen (liquid N_2, stored at -80°C). Resuspended cells were thawed and lysed by sonication and passaged through a French press system. Insoluble contents were removed by centrifugation (25,000 *g*, 20 min). The crude extracts were loaded onto Ni-NTA superflow resin columns (3 ml, Qiagen), 30 column volumes (CV) of TBS_W were applied for washing. Elution was achieved by 2.5 CV of TBS_E (TBS_W with 300 mM imidazole). Human rhinovirus 3C-protease (2% w/w) was mixed with eluted protein fractions and the reaction mixtures were dialyzed against 50 mM Tris pH 7.4 and 300 mM NaCl to remove the His$_6$-tags. Protease and uncleaved proteins were trapped by IMAC while cleaved protein was collected in the flow-through. Monomeric protein fractions were isolated by SEC on an Äkta explorer chromatography system using a HiLoad 16/60 Superdex 200 pg column and 10 mM Tris, pH 7.4 with 100 mM NaCl as running buffer. Proteins were concentrated (Amicon Ultra Centrifugal Filters, Merck Millipore) to a final concentration of approximately 20 mg/ml. Peptides (LifeTein) were

added with a 1.5-1.8 molar excess over protein. Protein preparations used to determine structures 5MFE, 5MFG, 5MFH and 5MFN were supplemented with 200 mM $CaCl_2$ to enhance the chance of crystal hits by enforcing Ca^{2+}-bridged dArmRP dimers (see below).

Crystallization and structure determination

Screening for initial conditions was done with commercial 96-well sparse-matrix screens (Hampton Research and Molecular Dimensions) at 4°C in a sitting-drop vapor diffusion set-up. For each condition the reservoir solution was mixed in three ratios with protein solution (1:1, 1:2 and 2:1 or 5:1). Based on the conditions identified in the initial screens, optimization screens (perpendicular pH- and precipitant gradients in 96-well plates) were applied to improve crystal quality. Supplementary Table ST1 summarizes the crystallization conditions as well as data collection and refinement statistics. Crystals were flash-frozen (liquid N_2) in mother liquor supplemented with 5-20 % glycerol or ethylene glycol. Diffraction data were collected on beam lines X06DA and X06SA at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) using a Pilatus detector system (Dectris Ltd). Data were processed using the XDS, XSCALE and XDSCONV software packages (Kabsch, 2010).

PHASER (McCoy et al., 2007) was used for molecular replacement to obtain initial phases, and previously determined structures of dArmRP were used as search models. Refinement was done using the programs REFMAC5 (Murshudov et al., 1999; Murshudov et al., 2011), BUSTER (Bricogne G. et al., 2016) and Phenix-Refine (Afonine et al., 2010a; Afonine et al., 2010b), followed by model building in COOT (Emsley and Cowtan, 2004; Emsley et al., 2010).

Curvature analysis

Coordinates of C^{α} atoms of neighboring repeat pairs (excluding flexible loop residue at position 23) were extracted from all experimental structures and used as an input into the make_symmdef_file.pl script from the Rosetta symmetry framework (DiMaio et al., 2011),

which calculated the corresponding curvature parameters and generated models with uniform curvature of the input repeats (symm.pdb-files). A model of an armadillo repeat with a bound (KR)-dipeptide was made by averaging the C^{α}-coordinates of 53 repeats from structures 5AEI, 5MFL and 5MFM that showed bidentate hydrogen bonds shorter than 3.1 Å from the N^{37} side chain to the peptide bond in the peptide backbone. This model was superimposed onto two consecutive repeats of symm.pdb-files, and the C^{α}(P/P+2) distance of the bound peptide residues was measured. We analyzed whether copies of dArmRP from the same AU exhibit the same curvature parameters by plotting C^{α}(P/P+2) distances of repeat pairs of all chains within the same AU followed by manual inspection.

To remove bias toward structures with many chains in the AU, all structures were classified and the curvature parameters were averaged accordingly. The following classification was used: (i) single chain in AU (no averaging) (ii) one population in AU, i.e. all molecules are highly similar (averaging among the same repeat pairs of all chains in the AU) and (iii) more than one population in AU (averaging among the same repeat pairs of chains of the same population). The grouping of all structures and the plots used to decide upon the category are shown in Supplementary Fig. S1, S2, S3 and S4, and are summarized in Supplementary Table ST2. In some cases when the classification was not obvious, additional criteria were considered (e.g., dArmRP chains that form Ca^{2+}-bridged dimers in the AU often display similar curvatures and were therefore considered as one population; or if crystal contacts differ largely between chains (PDB ID: 5MFI) they were treated as different populations, even if they show similar parameters; or if pairs of chains are practically identical, while there is a small difference between them, they were treated as two populations, PDB ID: 4V3Q and 4V3O). If the classification was still doubtful (PDB ID: 4DBA and 5MFH), chains were considered as one population; for details, see Supplementary Figs. S1-S4.

Statistical analysis was done in GraphPad Prism 6, a p-value of 0.05 was regarded as significant. Comparisons between two populations were evaluated with a Student's t-test (if an F-test showed no difference in variances of populations) or a Mann-Whitney test (if variances were different). Comparisons between more than two populations were done by one-way ANOVA (same variances in populations) or Kruskal-Wallis tests (different variances).

Single-molecule FRET experiments

For fluorescence labeling of $Y_{III}M_5A_{II}$ with a donor-acceptor dye pair, amino acid exchanges D51C and S273C were introduced by site-directed mutagenesis. The protein was expressed and purified by IMAC as describe above, but 1 mM β-mercaptoethanol (Pierce) was added to all buffers to keep the Cys residues reduced. SEC was conducted in PBS containing 1 mM DTT, followed by anion exchange chromatography in 20 mM NaH_2PO_4/Na_2HPO_4, pH 7.2 on a MonoQ 5/50 GL column. For elution, a gradient from 0 to 1000 mM NaCl was applied. For labeling, maleimide derivatives of Alexa 488 and 594 were used. These commercially available fluorophores contain a C5 linker, which ensures rotational averaging of the dye orientations, as well as optimal quantum yields. 1-2 mg of protein were incubated with 0.7 molar equivalents of Alexa Fluor-488 (ThermoFisher Scientific) at 4 °C for 16 h in a total volume of 2 ml. Singly labeled $Y_{III}M_5A_{II}$ was separated from unlabeled and doubly labeled protein by anion exchange chromatography and incubated with a 1.5-fold excess of Alexa Fluor-594 at 4 °C for 16 h. Finally, donor-acceptor labeled protein was purified by anion exchange chromatography, flash-frozen and stored at -80 °C until further use.

Fluorescently labeled protein was observed on a confocal single-molecule instrument (MicroTime 200, PicoQuant GmbH) at 22 °C as described before (Aznauryan et al., 2013). Dyes were excited using pulsed interleaved excitation (PIE) (Müller et al., 2005) at a pulse repetition rate of 20 MHz. Light of a supercontinuum fiber laser (SC-450-4, Fianium Ltd.,

Southampton, UK) selected with a z582/15 band pass filter (Chroma Technology) was used to excite Alexa Fluor-594, and Alexa Fluor-488 was excited using an LDH-D-C-485 diode laser (PicoQuant GmbH). Fluorescence photons were collected by the objective, passed through a chromatic beam splitter (R405/488/594 Chroma Technology), focused onto a 100 μm pinhole, and separated according to polarization and wavelength using a polarizing beam splitter and two 585DCXR dichroic mirrors (Chroma Technology). Finally, photons were filtered (ET525/50M or HQ650/100 band-pass filters, ChromaTechnology) and detected by four avalanche photodiodes (SPCMAQR-15, PerkinElmer Inc., Wellesley, MA). All measurements were conducted in phosphate-buffered saline (PBS) supplemented with 0.001 % (w/v) Tween-20 (Pierce) to minimize surface adhesion and 143 mM β-mercaptoethanol (Sigma-Aldrich) for photoprotection at a protein concentration of approximately 50 pM. The total acquisition time was 60 min per measurement.

Photon recordings were binned into 1 ms intervals, and fluorescence bursts caused by labeled proteins diffusing through the confocal volume were identified using a threshold of 30 photons per bin. Photon counts were corrected for background, differences in donor/acceptor quantum yields, differences in detection efficiencies, spectral crosstalk, as well as direct acceptor excitation (Schuler, 2007). For each photon burst, the stoichiometry ratio, *S*, and the transfer efficiency, *E*, were calculated as $S = \frac{n_{\mathrm{tot,Dex}}}{n_{\mathrm{tot,Dex}} + n_{\mathrm{tot,Aex}}}$ and $E = \frac{n_{\mathrm{A}}}{n_{\mathrm{A}} + n_{\mathrm{D}}}$, respectively. Here, $n_{\mathrm{tot,Dex}}$ and $n_{\mathrm{tot,Aex}}$ correspond to the corrected total numbers of photons emitted following donor or acceptor excitation, respectively, and n_{D} and n_{A} denote the corrected donor and acceptor photon counts following donor excitation (Müller et al., 2005). Bursts with $S < 0.6$ were used to generate transfer efficiency histograms. Probability distribution analysis (PDA) was used to analyze transfer efficiency histograms (Antonik et al., 2006). PDA assumes an intrinsic distribution of experimental transfer efficiencies, which is further broadened by shot noise. Shot-noise broadening is modeled using the photon statistics of the experimental burst-

size distribution. Here we applied PDA assuming a very narrow intrinsic transfer efficiency distribution to demonstrate that the width of the histogram peaks is dominated by shot-noise. Geometrically accessible volumes of dyes were calculated for chain B of 5AEI and chain C of 5MFC (PDB IDs), which served as model structures for the ligand-bound and ligand-free state, respectively. All calculations were performed using the freely available "FRET Positioning and Screening" software, in which Alexa 488 and 594 were parameterized as described earlier (Kalinin et al., 2012). To take into account the rapid fluctuations in interdye distance that occur on timescales similar to the excited state lifetime of the donor, we modeled the dynamics as diffusive motion in the potential of mean force corresponding to the interdye distance distribution from the accessible volume calculations using the diffusion coefficient of the free dyes (Best et al., 2015).

Results

Crystal structures serving as the basis of the curvature analysis

Here we are analyzing the curvature of all previously published structures of dArmRP (n=13), as well as another 14 previously unpublished structures. We expect that this large dataset will allow us to estimate the possible curvatures that can be adopted by dArmRP better than single structures that are possibly influenced by crystal packing. Overall, these 27 structures represent two different design approaches (consensus (Alfarano et al., 2012; Madhurantakam et al., 2012; Parmeggiani et al., 2008) and computational (Reichen et al., 2016b)), cover dArmRPs from 3 to 6 internal repeats, represent several different capping and internal repeat types, and are either apo structures or complexed with different peptide ligands. The structures were obtained in various space groups with 1-8 dArmRP molecules per asymmetric unit (AU) and resolutions ranging from 1.3-2.8 Å. An overview of all structures is given in Table 1, and the primary dArmRP sequences are shown in Fig. 2. Detailed discussions of the

previously published structures are provided in the original publications (Table 1). New structures will not be discussed in-depth, although interesting features will be highlighted.

A recurring arrangement in the crystal structures of consensus dArmRP is the formation of "Ca^{2+}-bridged dimers", where Ca^{2+} ions are complexed between two repeats of neighboring chains, thereby forming an antiparallel dimer (Hansen et al., 2017; Reichen et al., 2016a). This motif is found in roughly half of the consensus dArmRP structures but never in computational dArmRPs (Table 1). Different arrangements of Ca^{2+}-bridged dimers exist depending on which repeats of two dArmRP are paired by complexing the same Ca^{2+} ion. Most often a shift of 1 (i.e., one repeat) is observed (Fig. 3a, in structures 4V3Q, 4PLS, 5AEI, 5MFE, 5MFG, 5MFH, 5MFL and 5MFO). A shift of 0 is found in structures 5MFD and 5MFN while a shift of 2 is found in structures 4V3O and 5MFN.

5MFF forms tight parallel dArmRP dimers that are not mediated by Ca^{2+} ions; the stoichiometry with the $(RR)_5$ peptide is one peptide per two dArmRP. This is a consequence of the tight packing of the dArmRP dimers in this lattice, since it is sterically impossible to fit two peptides into the small channel between the two dArmRPs. In size exclusion chromatography experiments with a multi-angle light scattering detector (SEC-MALS), dimer formation by addition of $(RR)_5$ was not detected (data not shown). Hence we believe the 2:1 stoichiometry is a crystallization artefact and not present in solution. In this conformation the peptide side chains are involved in cation-π stacking interactions with W^{33} and salt bridges with E^{30} from both dArmRP chains of the dimer (Fig. 3b).

5MFG and 4PLS form a similar dArmRP tetramer; domain swapping events of the capping repeats connect two dArmRP such that the exposed internal repeats stack together to form one apparent dArmRP with ten internal repeats. The tetramer consists of two such dimers connected by Ca^{2+} ions (Reichen et al., 2014b) (Fig 3c). This domain swapping event is different from the one observed with the Y_{II} cap, where the cap was exchanged between two

neighboring chains (4DBA and 4DB8) (Madhurantakam et al., 2012). In 5MFG a peptide ligand is visible: interestingly, even though an octa-arginine peptide ($(RR)_4$) was added to the crystallization mixture, ten residues could be fitted to the electron density. This is probably due to multiple binding registers that are occupied within the crystal. Also the peptide deviates significantly from the expected canonical binding mode, as it is possible to only bind pockets formed by adjacent W^{33} residues that are in close proximity in this tetrameric arrangement. The pockets formed by W^{33} were shown to contribute higher binding energies to the peptide than interactions with other dArmRP residues (Hansen et al., 2016).

5MFI and 5MFJ crystallized in the same lattice, even though 5MFI was co-crystallized with $(KR)_4$ while it was $(KR)_5$in the case of 5MFJ. The lattice is made from parallel dArmRP dimers that form long tubes with symmetry-related molecules. The binding grooves of dimerized dArmRP are located in close proximity, and this feature is even extended to symmetry-related molecules. Therefore, the binding register of the peptides is not well defined, since most probably no completely homogenous complexes are formed throughout the crystals. In the models, two peptide stretches were built that resemble the clearest electron density, but also other registers are conceivable and the identity of the side chains (Lys or Arg) cannot always be unambiguously determined (Fig. 3d).

In 5MFL and 5MFM a novel strategy was used to enforce complex formation with the peptide, as $(KR)_5$ was directly fused via $(G_3S)_n$ linkers to either the N- (5MFL) or the C-terminus (5MFM) (Fig. 3e, sequences in Fig. 2). Both structures show canonical binding along the whole binding groove in each chain of the AU. In 5MFM electron density for one more repeat than present in the dArmRP is visible for chains F and C. We believe that this stems from inhomogeneous packing within the crystal, where these two dArmRP can be shifted by one repeat relative to the other chains in the AU. Only faint electron density for the terminal repeats is visible, since they do not have full occupancy. The N-cap (which had the

better defined electron density) was modeled within a full-length dArmRP and the faint additional density was modeled by an additional C-cap. Also for the peptide ligands of 5MFM additional electron density is observed in most chains. Hence peptides were modeled as two chains of alternative conformations that are shifted by one or two (KR) repeats relative to each other with a total occupancy of 1. Here the electron density clearly indicates the occurrence of several binding registers of the peptide along the dArmRP binding groove. Such multiple binding modes were proposed to contribute to the high affinity of dArmRP (Hansen et al., 2016).

Curvature analysis

Curvature parameters were calculated from neighboring armadillo repeat pairs of all structures listed in Table 1 (see Material and Methods, Fig. 1). A total of 329 repeat pairs was analyzed. All raw data of this analysis is supplied in the Supplementary raw data table.

Analysis of dArmRPs from the same AU showed that in some structures all chains in the AU adopt the same curvature, while in others up to 4 populations with unique curvature parameters are found per AU, demonstrating that dArmRPs possess some inherent flexibility (Supplementary Fig. 3 and 4 and Supplementary Table ST2).

We then generated a normalized dataset to remove bias towards structures with many chains per AU while retaining the information of chains that adopt different conformations within the same AU. This was achieved by averaging parameters of dArmRP from the same AU exhibiting identical curvatures in a repeat pair-wise fashion (see Material and Methods and Supplementary Material for further details) and thus counting them only once. After averaging, the curated data set consisted of 151 sets of curvature parameters (Supplementary raw data table).

The mean rise in this datasets is 7.34 ± 0.72 Å (mean value ± standard deviation) with minimal and maximal values of 5.95 and 10.45 Å, respectively. The radius r_A (to the CoM)

has a mean value of 15.26 ± 2.25 Å (7.69-20.25 Å), while the angle 2Ω adopts values ranging from 23.0° to 32.0° with a mean of 27.7 ± 1.8°. These parameters are not independent of each other: the angle-radius distribution of superhelical repeat proteins follows a hyperbolic shape (Brunette et al., 2015). In our dataset 2Ω angles plotted as a function of radii show a curve with a negative slope. This is in agreement with the hyperbolic distribution since only a small angle-range is sampled and the negative slope is consistent with the fact that dArmRPs form right-handed super helices. The rise is not directly dependent on the other parameters. For α-helical repeat proteins, values below 16 Å are adopted, since a larger rise does not allow the formation of a continuous hydrophobic core (Brunette et al., 2015). The most informative parameter is the C^{α}(P/P+2) distance, as it directly assesses whether modular peptide binding over long stretches is possible with the respective curvature.

The curvature parameters of the consensus and computational dArmRPs can be clearly distinguished; the computational design is characterized by a significantly higher rise and angle but a smaller radius (h = 7.95 ± 0.55 Å, 2Ω = 30.0 ± 1.1°, r_A = 12.44 ± 1.58 Å). The parameters of the consensus dArmRP show some pronounced outliers that can be explained by domain swapping events in the parental structures. Repeat pairs directly adjacent to domain-swapped Y_{II}-caps (in structures 4DBA and 4DB8) entail a pronounced enlargement of rise and C^{α}(P/P+2) distance. Domain swaps of the Y_{III}-cap (in structures 4PLS and 5MFG, chains B and C) do not lead to elevated rise or C^{α}(P/P+2) values. On the other hand the domain swap of the A_{II}-cap in structure 5MFG (chains A and D) leads to a very small rise and C^{α}(P/P+2) distance in adjacent repeat pairs (Supplementary Fig. S1 and S3). Since we consider that domain-swaps are crystallization artefacts, the respective curvature parameters from those structures were removed from all following analyses (dataset: cons-dds). In this dataset of consensus dArmRP, h adopts values of 7.12 ± 0.52 Å, 2Ω of 27.2 ± 1.4° and r_A of 16.16 ± 1.48 Å (Fig. 4a-d). Computational dArmRP form a rather tall and thin super helix. This results in C^{α}(P/P+2) distances of 8.03 ± 0.60 Å, which is incompatible with continuous

peptide binding as described previously (Reichen et al., 2016b). Consensus dArmRPs, on the other hand, form shorter but wider superhelices (as described for the optimal curvature model (Reichen et al., 2016b)). The mean C^{α}(P/P+2) values imply that the consensus dArmRPs are better suited to continuously bind peptides, though, the mean C^{α}(P/P+2) value of 7.35 ± 0.42 Å is still slightly above the optimal range.

To investigate whether peptide binding can induce a more optimal curvature we divided the cons-dds and computational dataset into four sub-datasets based on the ligand binding properties of the respective structure. Henceforth the sub-datasets are referred to as canonical (bidentate hydrogen bonds with N^{37} residues are observed along the whole dArmRP), partial canonical (bidentate bonds present in some instances), non-canonical (peptide does not interact with N^{37} residues and no 1:1 stoichiometry is present) and no ligand (no peptide ligand present in the structure). The grouping of the structures is given in Table 1. In consensus dArmRP, canonically bound peptides induce a significantly smaller C^{α}(P/P+2) value compared to partial canonical and structures without ligand. Structures with non-canonical binding exhibit C^{α}(P/P+2) values similar to the canonical structures. These data are derived from the structures 5MFF and 5MFG, in which the crystal lattices make it sterically impossible (5MFF) or energetically unfavorable (5MFG) to form canonical complexes, even though the curvature parameters appear to be suitable (see above and Fig. 5a). The C^{α}(P/P+2) distance has a mean of 7.04 ± 0.21 Å for the canonical, 7.03 ± 0.21 Å for the non-canonical, 7.43 ± 0.42 Å for the partial canonical and 7.41 ± 0.42 Å for the no ligand sub-dataset. Hence, canonically bound structures of consensus dArmRPs have C^{α}(P/P+2) values that are suitable for modular peptide binding. Furthermore, peptide binding seems to reduce protein-to-protein variation in curvature parameters, as the corresponding standard deviations are smaller compared to the ones of the partial canonical and no ligand sub-datasets.

For the computationally designed dArmRPs only structures with partial canonical binding or no ligand were determined. Their C^{α}(P/P+2) distances adopt values of 7.96 ± 0.48 Å and 8.15 ± 0.77 Å for the partial canonical and the no ligand sub-dataset, respectively. Not only are these values very similar, they are also considerably greater than the optimal range (Fig. 5b).

To investigate whether capping repeats influence the curvature of neighboring internal repeats, we grouped C^{α}(P/P+2) distances into N-cap neighboring, central and C-cap neighboring repeat pairs. For both design approaches, repeat pairs adjacent to caps display elevated C^{α}(P/P+2) values . But only for N-cap neighboring repeats of consensus dArmRP the C^{α}(P/P+2) distances are significantly larger than the C^{α}(P/P+2) values of central repeats (7.59 ± 0.40 Å and 7.21 ± 0.40 Å, respectively; Fig. 5c and 5d).

The formation of Ca^{2+}-bridged dimers leads to significantly lower C^{α}(P/P+2) values (7.27 ± 0.35 Å in Ca^{2+}-bridged dimers, 7.56 ± 0.51 Å without Ca^{2+}-bridged dimers). However, the entire canonical sub-dataset falls into the group with Ca^{2+}-bridged dimers. If the values of the canonical sub-dataset are removed, the difference between the groups is not significant anymore (7.34 ± 0.37 Å with Ca^{2+}-bridged dimers, 7.56 ± 0.51 Å without Ca^{2+}-bridged dimers). Thus, Ca^{2+}-bridged dimer formation might help with the adaptation of consensus dArmRP to the peptide by regularizing the arrangement, but the main contribution stems from the interaction with the peptide.

Single-molecule FRET experiments

As the analysis of all crystal structures revealed, the curvature of dArmRP can be influenced by the specific arrangement of a crystal lattice. We were interested in the behavior in solution which is the relevant environment for the final applications of dArmRP. Hence we wanted to probe the presence of conformational changes upon peptide binding unbiased by the context of the protein crystal. Therefore, we performed single-molecule Förster resonance energy transfer (FRET) spectroscopy of $Y_{III}M_5A_{II}$ in solution. Single-molecule FRET relies on the

distance-dependent energy transfer between a donor and an acceptor fluorophore (Selvin and Ha, 2008). For this purpose, cysteine residues were introduced at positions 51 and 273 in the consensus dArmRP sequence, followed by sequential coupling of maleimide-functionalized Alexa Fluor 488 (donor) and Alexa Fluor 594 (acceptor). Confocal single-molecule FRET experiments were performed at a protein concentration of approximately 50 pM in the presence and absence of 1 μM unlabeled $(KR)_5$, respectively. Individual FRET-labeled $Y_{III}M_5A_{II}$ molecules diffusing through the confocal volume were detected as bursts of fluorescence photons. Transfer efficiencies were calculated for each burst and binned in histograms. Since single molecules are measured, single-molecule FRET enables us to (i) resolve subpopulations that would otherwise be hidden in the ensemble average, and (ii) compare the width of transfer efficiency distributions to the value theoretically expected from photon shot noise.

The observed transfer efficiency histograms are depicted in Fig. 6a. In both cases, a single, well-defined peak is observed. Note that the widths of the peaks are dominated by the stochastic fluctuation of recorded numbers of donor and acceptor photons (shot noise). This is illustrated by calculating the theoretically expected shot noise broadening using probability distribution analysis (PDA) (Fig. 6a, black lines) (Antonik et al., 2006). This observation indicates that the distance fluctuations between donor and acceptor are faster than the burst duration, and only mean transfer efficiencies, $\langle E \rangle$, can be obtained from the histograms. We found an increase from $\langle E \rangle_{\text{exp}} = 0.21$ in the absence to $\langle E \rangle_{\text{exp}} = 0.31$ in the presence of $(KR)_5$, indicating a decrease in the average inter-dye distance.

To directly compare single-molecule FRET experiments and crystallographic results, we estimated the expected mean transfer efficiencies based on the atomic-resolution structures. Here, chain B of 5AEI served as a model for bound dArmRP, and chain C of 5MFC as a model of unbound dArmRP (Fig 7b). Since the linkers used to conjugate the dyes to the

proteins are long and flexible, a relatively large range of inter-dye distances can be sampled. To calculate the expected mean transfer efficiencies, we determined the distribution of interdye distances from the geometrically accessible volumes of both dyes (Fig. 6b) (Kalinin et al., 2012) and take into account the dynamics of inter-dye distance fluctuations (see Methods). This analysis yields an expected change in transfer efficiency from approximately 0.23 to 0.28 upon peptide binding, very similar to our experimentally determined values. In summary, the single-molecule FRET experiments thus confirm a small but significant distance change that occurs on the sub-nanometer scale upon canonical peptide-binding to dArmRPs.

Discussion

In this study, we show that dArmRPs adopt different curvature parameters in different crystal structures, sometimes even within the same AU. This demonstrates an intrinsic flexibility of the scaffold which was also observed for nArmRP (Huber et al., 1997) and importin-β, belonging to the closely related family of HEAT repeat proteins (Halder et al., 2015; Tauchert et al., 2016; Zachariae and Grubmüller, 2008). However, we expect the flexibility to be smaller in dArmRPs since the scaffold is more regular and stable compared to nArmRP (Madhurantakam et al., 2012; Parmeggiani et al., 2008; Reichen et al., 2016b).

The packing of the crystal lattice plays a major role for the observed curvature but it is very difficult to predict the specific influence of the lattice on the curvature of the protein, since multiple crystal contacts have to be taken into account that are hard to quantify and to compare between different lattices. Moreover, the differences among all structures are still small. The superposition of C^{α}s of chain B of 5AEI (consensus dArmRP with the lowest average C^{α}(P/P+2) value of 6.93 Å) and chain A of 5MFC (consensus dArmRP with the highest average C^{α}(P/P+2) value of 7.84 Å) results in a RMSD of 1.49 Å. Thus, for most scaffold design projects such differences would not be considered meaningful. However, for

the purpose of designing a modular peptide-binding scaffold, a value close to the optimum is important for long-range peptide binding. In solution, a range of curvatures around an average state will be adopted while in a crystal structure one or few possible states that are locked in the crystal lattice are observed. Hence, it is not possible to characterize the curvature that is adapted in solution by a single crystal structure. However, we deem that the large set of structures used herein samples a sufficiently large conformational space in order to estimate the average curvature in solution.

Our analysis shows that the two design approaches – consensus and computational – can be distinguished by all the parameters used to describe the curvature. It is remarkable that the consensus dArmRP exhibits a curvature that is better suited to bind peptides continuously than the computational dArmRP, since these were designed by using curvature restraints, while no such restraints were used during the consensus design. It should be noted, however, that several mutations had to be introduced to the computational dArmRPs in order to obtain diffracting crystals. Hence, it remains unclear whether the original computational designs would adopt the desired curvature (Reichen et al., 2016b) as they could not be crystallized.

Consensus dArmRP have a C^{α}(P/P+2) value of 7.35 ± 0.42 Å which is slightly above the optimal value (6.7-7.0 Å). However, peptide binding can induce curvatures that are in agreement with modular binding as illustrated by structures 5AEI and 5MFM (average C^{α}(P/P+2) value of 6.94 Å and 6.98 Å, respectively) (Fig. 5). The induction of an optimal curvature by peptide binding is corroborated by the fact that none of the ligand-free structures exhibits a similar geometry and by the changes on peptide binding observed in single-molecule FRET measurements (see below).

Forces of the crystal lattice are an additional parameter influencing the curvature. These appear to be dominant over the binding energy of the peptide, since in a number of structures no ligand is found (even though it was added to the crystallization mix), or only partial

canonical binding is observed. We believe that in many cases the lattice enforces a curvature that is not adequate for modular binding or only allows modular binding of short peptide stretches. In addition, peptide binding might be affected by additives used to obtain crystals that lower the affinity of the interaction (e.g. high salt or extreme pH values) (Hansen et al., 2016). In some cases, non-canonical binding is observed even though a suitable curvature is adopted. Here steric or energetic effects are responsible for the deviation from the expected binding mode (e.g., 5MFF and 5MFG, see Results).

The curvature of computational dArmRP was analyzed previously for 4D4E and 4D49 (Reichen et al., 2016b). Here we extend this analysis with additional structures (5MFB, 5MFI, 5MFJ and 5MFK) to obtain a more representative dataset. The findings from the previous study that computational dArmRP possess curvatures which only allow limited modular peptide binding are confirmed by the additional structures analyzed in this study.

Capping repeats are known to be important for the stability of repeat proteins in general; in the case of dArmRP during the consensus design the caps were reengineered several times to improve the stability (Alfarano et al., 2012; Forrer et al., 2004; Interlandi et al., 2008; Madhurantakam et al., 2012; Wetzel et al., 2008). We found that neighboring repeat pairs of caps, especially of the N-cap from consensus dArmRP, possess slightly higher C^{α}(P/P+2) values (Fig. 5c and 5d). This might indicate that the interface between these repeats is not optimal. However, this issue can be easily addressed by introducing a "spacer repeat" between caps and binding repeats. Such an approach has been shown to even improve protein stability in library construction (Varadamsetty et al., 2012).

The analysis of the crystal structures suggests that consensus dArmRPs undergo a change in C^{α}(P/P+2) distances upon $(KR)_5$ peptide binding. Single-molecule FRET experiments of donor/acceptor labeled $Y_{III}M_5A_{II}$ performed in the presence and absence of $(KR)_5$ show an increase in the mean transfer efficiency from 0.21 to 0.31 upon peptide binding. These values

are in good agreement with those calculated from the crystal structures, taking the dynamics of the dye linker into account. Taken together, single-molecule FRET experiments demonstrate that the small conformational change of consensus dArmRPs observed in the crystal structures is also present in dArmRP in free solution.

In summary, our analysis shows that the consensus dArmRP scaffold is suitable as a basis for the development of modular peptide binders. An inherent flexibility allows the scaffold to adopt a suitable curvature for modular peptide binding. Further engineering of the scaffold could be envisaged to optimize the curvature also in the unbound state (i.e. decrease C^{α}(P/P+2) distance) or possibly even by increasing the flexibility of the scaffold. This might increase the affinity towards peptides, even though this does not appear to be a pressing problem, as picomolar affinities can already be reached with the present design (Hansen et al., 2016). Furthermore, we could spotlight that in design approaches, where very high accuracy of the scaffold design is important, protein crystallization can introduce small but significant deviations from possible geometries in solution since it does not reflect the inherent flexibility of scaffolds and lattice forces can be dominant over induced changes by ligand-binding.

Acknowledgements

We would like to thank Céline Stutz-Ducommun and Beat Blattmann from the UZH Protein Crystallization Center for help with crystallization experiments and the staff of beamlines X06DA and X06SA at the Swiss Light Source (Paul Scherrer Institut, Würenlingen, Switzerland) for technical support. This work was supported by the Swiss National Science foundation (grant number CRSI_141832/1 to A.P. and 31003A_152839 to B.S.), COST Action BM1405 (Non-globular proteins) and the Forschungskredit of the University of Zurich (grant numbers FK-13-028 to S.H. and FK-16-018 to P.E.)

References

Afonine, P.V., Mustyakimov, M., Grosse-Kunstleve, R.W., Moriarty, N.W., Langan, P., Adams, P.D., 2010a. Joint X-ray and neutron refinement with phenix.refine. Acta Crystallogr. Sect. D. Biol. Crystallogr. 66, 1153-1163.

Afonine, P.V., Grosse-Kunstleve, R.W., Chen, V.B., Headd, J.J., Moriarty, N.W., Richardson, J.S., Richardson, D.C., Urzhumtsev, A., Zwart, P.H., Adams, P.D., 2010b. phenix.model_vs_data: a high-level tool for the calculation of crystallographic model and data statistics. J Appl Crystallogr 43, 669-676.

Alfarano, P., Varadamsetty, G., Ewald, C., Parmeggiani, F., Pellarin, R., Zerbe, O., Plückthun, A., Caflisch, A., 2012. Optimization of designed armadillo repeat proteins by molecular dynamics simulations and NMR spectroscopy. Protein Sci. 21, 1298-1314.

Antonik, M., Felekyan, S., Gaiduk, A., Seidel, C.A.M., 2006. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. J. Phys. Chem. B 110, 6970-6978.

Aznauryan, M., Nettels, D., Holla, A., Hofmann, H., Schuler, B., 2013. Single-molecule spectroscopy of cold denaturation and the temperature-induced collapse of unfolded proteins. J. Am. Chem. Soc. 135, 14040-14043.

Best, R.B., Hofmann, H., Nettels, D., Schuler, B., 2015. Quantitative Interpretation of FRET Experiments via Molecular Simulation: Force Field and Validation. Biophys. J. 108, 2721-2731.

Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P., Paciorek W., Roversi P, Sharff A., Smart O.S., Vonrhein C., T.O., W., 2016. BUSTER version 2.10.2 Cambridge, United Kingdom: Global Phasing Ltd.

Brunette, T.J., Parmeggiani, F., Huang, P.S., Habha, G.B., Ekiert, D.C., Tsutakawa, S.E., Hura, G.L., Tainer, J.A., Baker, D., 2015. Exploring the repeat protein universe through computational protein design. Nature 528, 580-584.

Conti, E., Kuriyan, J., 2000. Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha. Structure 8, 329-338.

Conti, E., Uy, M., Leighton, L., Blobel, G., Kuriyan, J., 1998. Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. Cell 94, 193-204.

DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D., André, I., 2011. Modeling symmetric macromolecular structures in Rosetta3. PLoS One 6, e20450.

Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. Sect. D. Biol. Crystallogr. 60, 2126-2132.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallographica Section D-Biological Crystallography 66, 486-501.

Forrer, P., Binz, H.K., Stumpp, M.T., Plückthun, A., 2004. Consensus design of repeat proteins. ChemBioChem 5, 183-189.

Graham, T.A., Weaver, C., Mao, F., Kimelman, D., Xu, W.Q., 2000. Crystal structure of a beta-catenin/Tcf complex. Cell 103, 885-896.

Gul, I.S., Hulpiau, P., Saeys, Y., van Roy, F., 2016. Metazoan evolution of the armadillo repeat superfamily. Cell. Mol. Life Sci., 1-17.

Halder, K., Dölker, N., Van, Q., Gregor, I., Dickmanns, A., Baade, I., Kehlenbach, R.H., Ficner, R., Enderlein, J., Grubmüller, H., Neumann, H., 2015. MD simulations and FRET reveal an environment-sensitive conformational plasticity of importin-beta. Biophys. J. 109, 277-286.

Hansen, S., Kiefer, J.D., Madhurantakam, C., Mitt, P.R.E., Plückthun, A., 2017. Structures of designed armadillo repeat proteins binding to peptides fused to globular domains. Protein Sci. in press.

Hansen, S., Tremmel, D., Madhurantakam, C., Reichen, C., Mittl, P.R., Plückthun, A., 2016. Structure and energetic contributions of a designed modular peptide-binding protein with picomolar affinity. J. Am. Chem. Soc. 138, 3526-3532.

Huber, A.H., Nelson, W.J., Weis, W.I., 1997. Three-dimensional structure of the armadillo repeat region of beta-catenin. Cell 90, 871-882.

Interlandi, G., Wetzel, S.K., Settanni, G., Plückthun, A., Caflisch, A., 2008. Characterization and further stabilization of designed ankyrin repeat proteins by combining molecular dynamics simulations and experiments. J. Mol. Biol. 375, 837-854.

Kabsch, W., 2010. XDS. Acta Crystallogr. Sect. D. Biol. Crystallogr. 66, 125-132.

Kalinin, S., Peulen, T., Sindbert, S., Rothwell, P.J., Berger, S., Restle, T., Goody, R.S., Gohlke, H., Seidel, C.A.M., 2012. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. Nat. Methods 9, 1218-U1129.

Kobe, B., Kajava, A.V., 2000. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. Trends Biochem. Sci 25, 509-515.

Madhurantakam, C., Varadamsetty, G., Grütter, M.G., Plückthun, A., Mittl, P.R., 2012. Structure-based optimization of designed Armadillo-repeat proteins. Protein Sci. 21, 1015-1028.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658-674.

Müller, B.K., Zaychikov, E., Brauchle, C., Lamb, D.C., 2005. Pulsed interleaved excitation. Biophys. J. 89, 3508-3522.

Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S., Dodson, E.J., 1999. Efficient anisotropic refinement of macromolecular structures using FFT. Acta Crystallogr. Sect. D. Biol. Crystallogr. 55, 247-255.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A., 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallographica Section D-Biological Crystallography 67, 355-367.

Nüsslein-Volhard, C., Wieschaus, E., 1980. Mutations affecting segment number and polarity in Drosophila. Nature 287, 795-801.

Park, K., Shen, B.W., Parmeggiani, F., Huang, P.S., Stoddard, B.L., Baker, D., 2015. Control of repeat-protein curvature by computational protein design. Nat. Struct. Mol. Biol. 22, 167-174.

Parmeggiani, F., Pellarin, R., Larsen, A.P., Varadamsetty, G., Stumpp, M.T., Zerbe, O., Caflisch, A., Plückthun, A., 2008. Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. J. Mol. Biol. 376, 1282-1304.

Reichen, C., Hansen, S., Plückthun, A., 2014a. Modular peptide binding: From a comparison of natural binders to designed armadillo repeat proteins. J. Struct. Biol. 185, 147-162.

Reichen, C., Madhurantakam, C., Plückthun, A., Mittl, P.R.E., 2014b. Crystal structures of designed armadillo repeat proteins: Implications of construct design and crystallization conditions on overall structure. Protein Sci. 23, 1572-1583.

Reichen, C., Madhurantakam, C., Hansen, S., Grütter, M.G., Plückthun, A., Mittl, P.R.E., 2016a. Structures of designed armadillo-repeat proteins show propagation of inter-repeat interface effects. Acta Crystallogr. Sect. D. Biol. Crystallogr. 72, 168-175.

Reichen, C., Hansen, S., Forzani, C., Honegger, A., Fleishman, S.J., Zhou, T., Parmeggiani, F., Ernst, P., Madhurantakam, C., Ewald, C., Mittl, P.R., Zerbe, O., Baker, D., Caflisch,

A., Plückthun, A., 2016b. Computationally designed armadillo repeat proteins for modular peptide recognition. J. Mol. Biol. 428, 4467-4489.
Schuler, B., 2007. Application of single molecule Forster resonance energy transfer to protein folding. Methods Mol. Biol. 350, 115-138.
Selvin, P.R., Ha, T., 2008. Single-Molecule Techniques: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.
Tauchert, M.J., Hemonnot, C., Neumann, P., Koster, S., Ficner, R., Dickmanns, A., 2016. Impact of the crystallization condition on importin-beta conformation. Acta Crystallogr. Sect. D. Biol. Crystallogr. 72, 705-717.
Tewari, R., Bailes, E., Bunting, K.A., Coates, J.C., 2010. Armadillo-repeat protein functions: questions for little creatures. Trends Cell Biol. 20, 470-481.
Varadamsetty, G., Tremmel, D., Hansen, S., Parmeggiani, F., Plückthun, A., 2012. Designed Armadillo repeat proteins: library generation, characterization and selection of peptide binders with high specificity. J. Mol. Biol. 424, 68-87.
Wetzel, S.K., Settanni, G., Kenig, M., Binz, H.K., Plückthun, A., 2008. Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins. J. Mol. Biol. 376, 241-257.
Zachariae, U., Grubmüller, H., 2008. Importin-beta: Structural and dynamic determinants of a molecular spring. Structure 16, 906-915.

Figure Captions

Fig. 1. Modular peptide binding. (a) Model of a continuous modular peptide binder made by propagation of internal repeats 1-5 of 5MFL (chain B, blue) with the bound peptide (red). (b) Schematic illustration of two intertwined helices with parameters to describe the curvature.

Fig. 2. Sequence alignment of all dArmRP constructs used in this study. The recognition site of 3C-protease is shown in yellow, the scissile bond is indicated by arrows. Positions of helices H1, H2 and H3 are indicated by green, blue and red rectangles, respectively. Repeats Y_{II}, Y_{III}, M, M' and M'', A_{II} and A_{III} are consensus-based; repeats Dq, Dq.V1 Dq.V2, C_{PAF} and CqI originate from the computational design approach. Sequence differences within consensus or computational designs are highlighted in red, differences between consensus and computational designs are shown by a grey background.

Fig. 3. Selected features of dArmRP structures. (a) Ca^{2+}-bridged dimer. *Top*: Overview of the arrangement of a typical Ca^{2+}-bridged dimer (chains A (blue) and B (red) with Ca^{2+} ions (yellow) from 5MFE) with a shift of 1; inset: close-up view of Ca^{2+}-complexing residues ($P^{23}N^{24}E^{25}$ in both monomers, respectively); *bottom*: scheme of Ca^{2+}-bridged dimers with different shifts found in experimental structures. (b) Parallel dimer formed in 5MFF. *Top*: Overview, chain A in blue, chain B in red, peptide in yellow; *bottom*: close-up view of peptide interactions (chain B and additional peptide residues removed for clarity). (c) Schematic display of domain swapping events. *Left*: domain-swapped dimer found with Y_{II} cap; *right*: domain-swapped tetramer found with Y_{III} cap and Ca^{2+} ions (yellow). (d) Crystal lattice of 5MFI and 5MFJ (5MFJ is shown here). Chain A in blue; chain B in red; peptides in yellow and green, respectively; symmetry-related chain A as blue surface; symmetry-related peptide in purple. (e) Schematic display of C- and N-terminal peptide fusions.

Fig. 4. Comparison of curvature parameters of consensus (blue and red) and computational (green) dArmRP; consensus: dataset containing all curvature parameters of consensus dArmRP; cons-dds: dataset of consensus dArmRP where curvature parameters of repeat pairs adjacent to domain-swaps were removed. (a) C^{α}(P/P+2) distance. (b) rise h. (c) angle 2Ω. (d) radius r_A. Boxes indicate 25-75 percentile with median value as horizontal line, whiskers indicate 5-95 percentile, values outside 5-95 percentile are shown as individual symbols. Significance levels are indicated by black brackets (****: $p<0.0001$).

Fig. 5. Influence of peptide binding and capping repeats on the curvature of dArmRP. (a) Influence of peptide binding on consensus dArmRP. (b) Influence of peptide binding on computational dArmRP. (c) Influence of proximity to capping repeats on consensus dArmRP. (d) Influence of proximity to capping repeats on computational dArmRP. Boxes indicate 25-75 percentiles with median value as horizontal line, whiskers indicate 5-95 percentiles, values outside 5-95 percentiles are shown as individual symbols. Significance levels are indicated by black brackets (**: $p<0.01$; ***: $p<0.001$).

Fig. 6. Influence of peptide binding on the conformation of $Y_{III}M_5A_{II}$ as studied by single-molecule FRET. (a) Transfer efficiency histograms of $Y_{III}M_5A_{II}$ labeled at positions 51 and 273 recorded in the absence (red) and presence (blue) of 1 μM $(KR)_5$. Measured transfer efficiency histograms (dark red and dark blue) are compared to peaks (pastel with black outline) calculated from probability distribution analysis (PDA) assuming that the peak widths are shot noise-limited. Vertical dashed lines indicate the mean transfer efficiencies used for the analysis. (b) Ribbon representations of $Y_{III}M_5A_{II}$ in the unbound (red, chain C of 5MFC) and the bound state (blue, chain B of 5AEI). The sterically accessible volumes of the donor (Alexa Fluor 488) and acceptor (Alexa Fluor 594) dyes are depicted as blue and red clouds.

Figure 1

Click here to download high resolution image

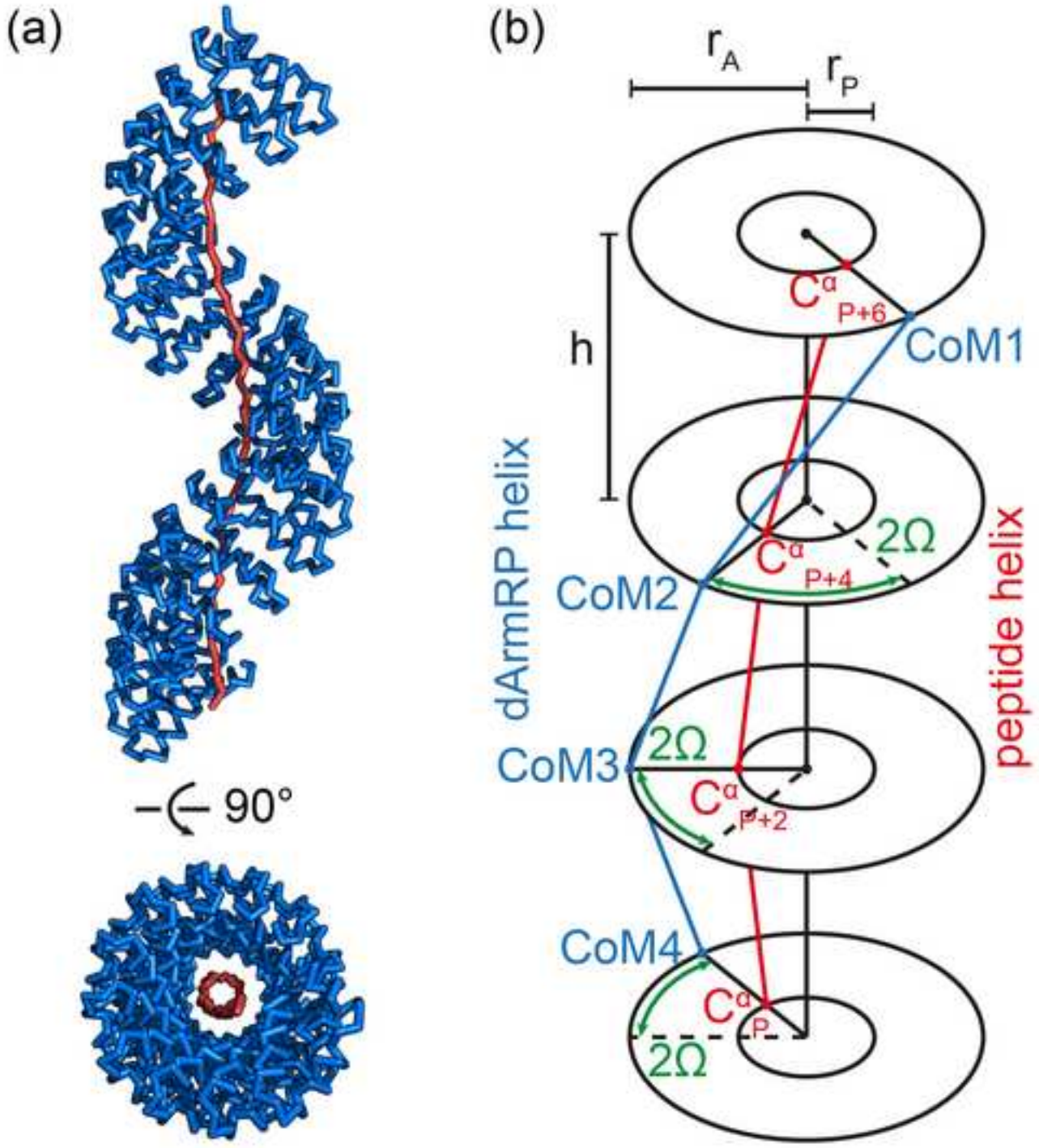

Figure 2

Click here to download high resolution image

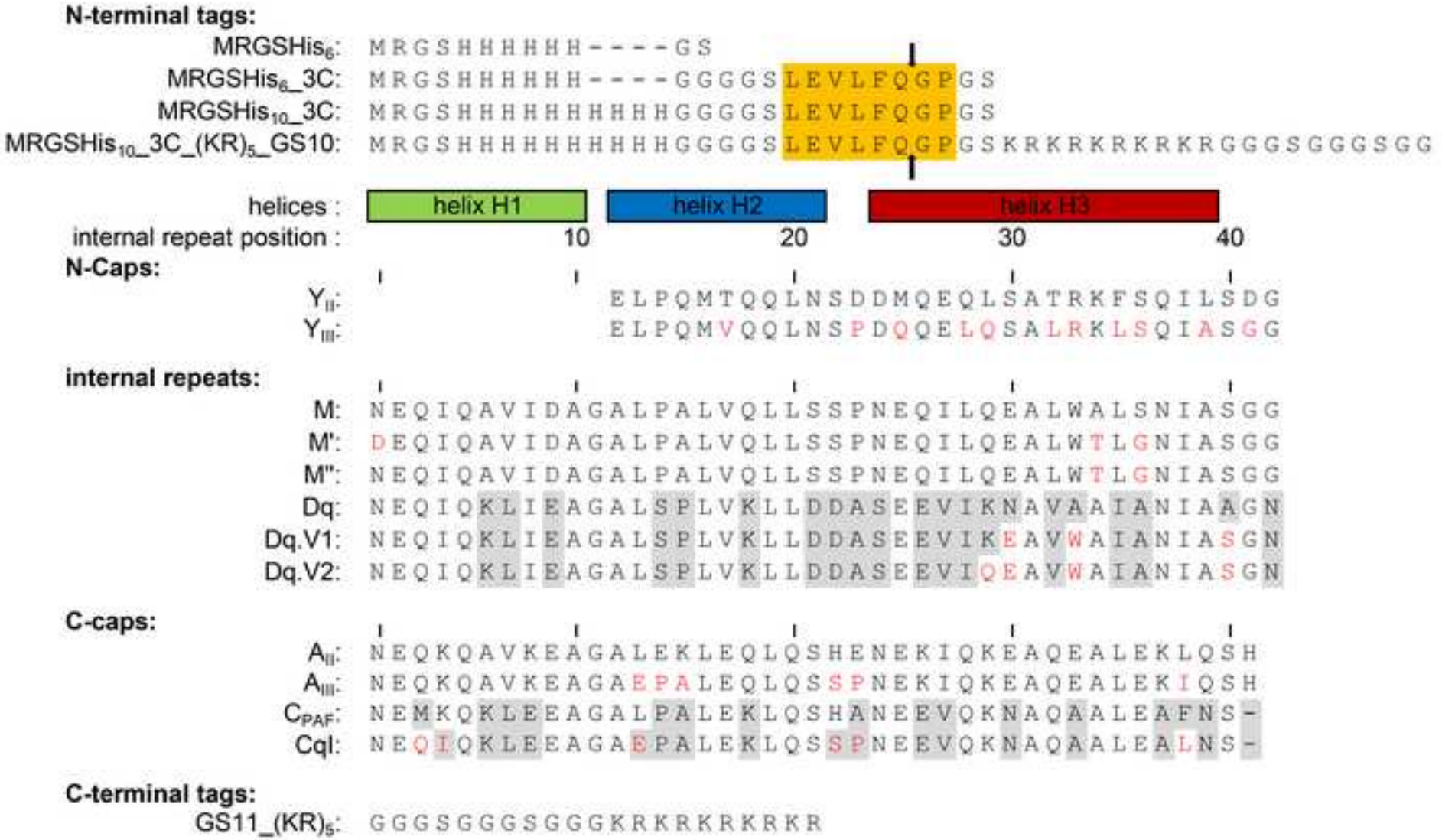

Figure 3

Click here to download high resolution image

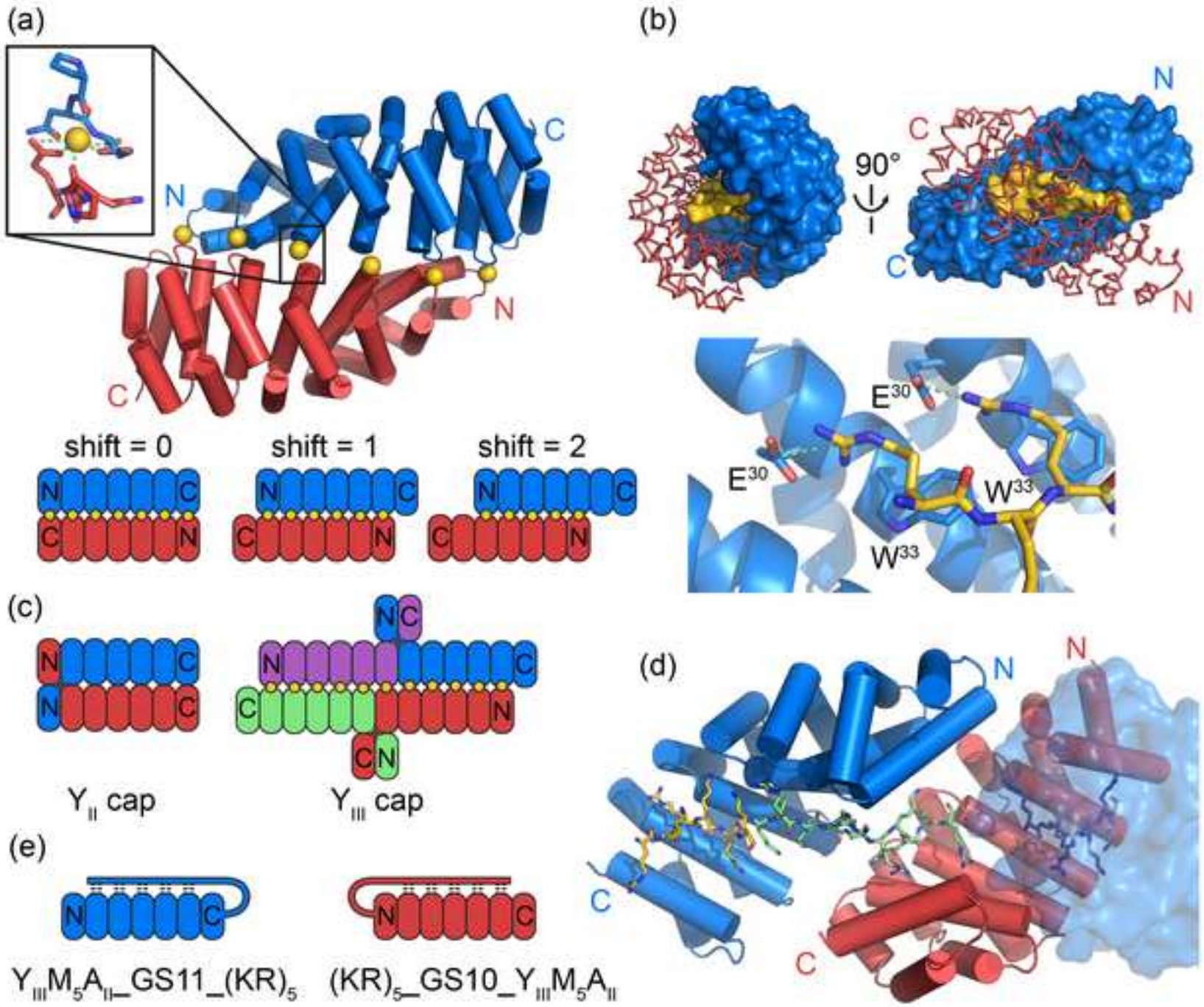

Figure 4

Click here to download high resolution image

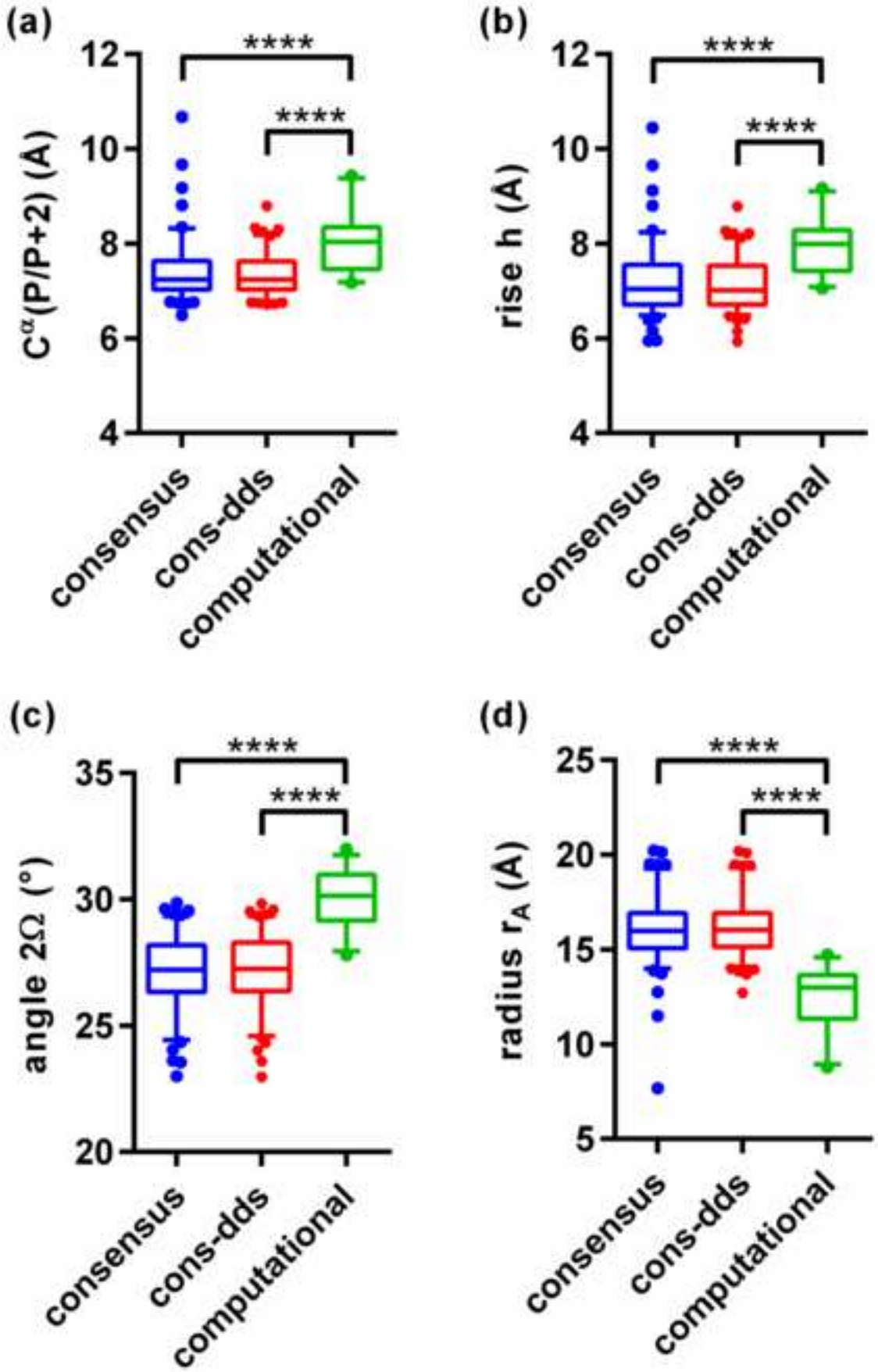

Figure 5
Click here to download high resolution image

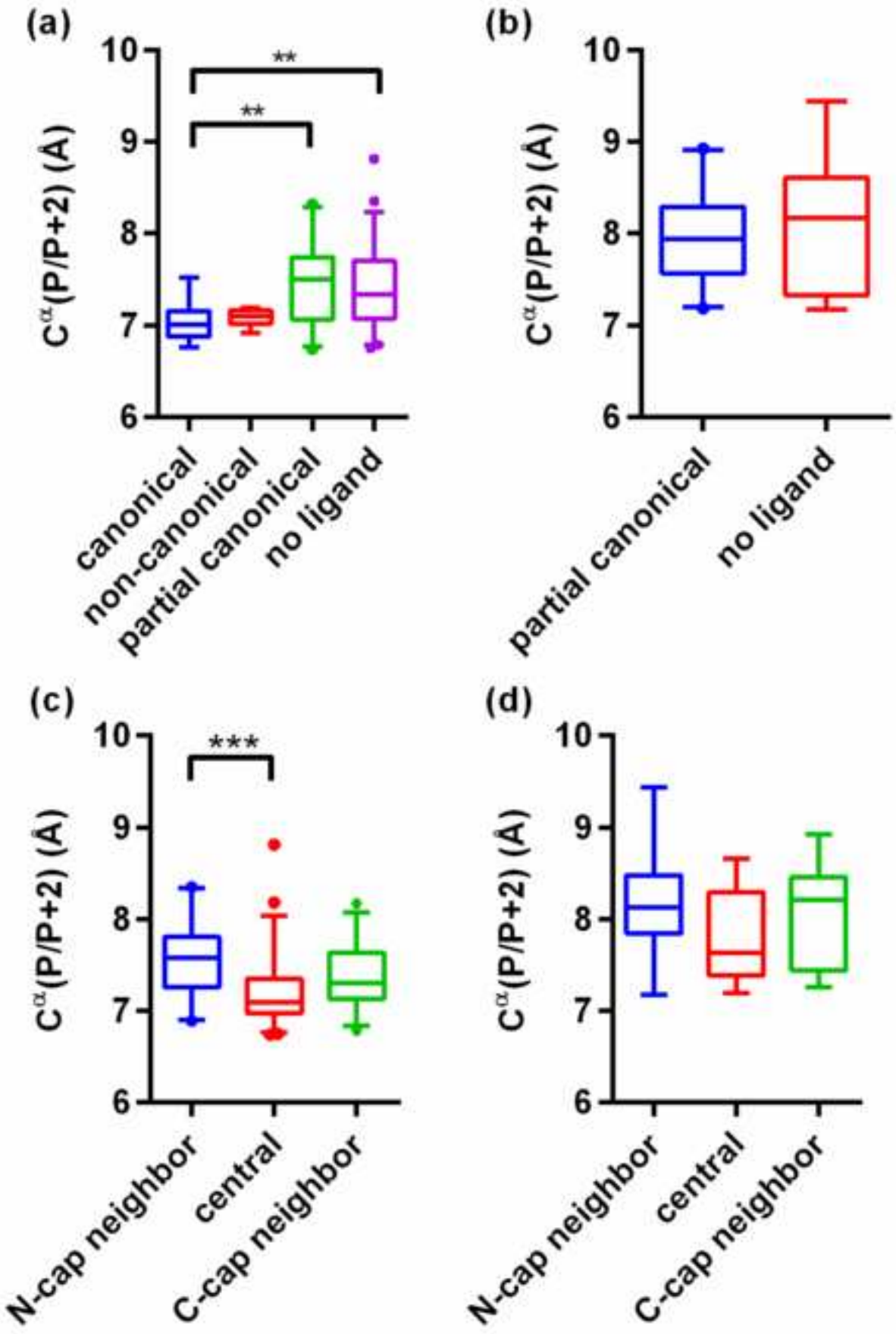

Figure 6
Click here to download high resolution image

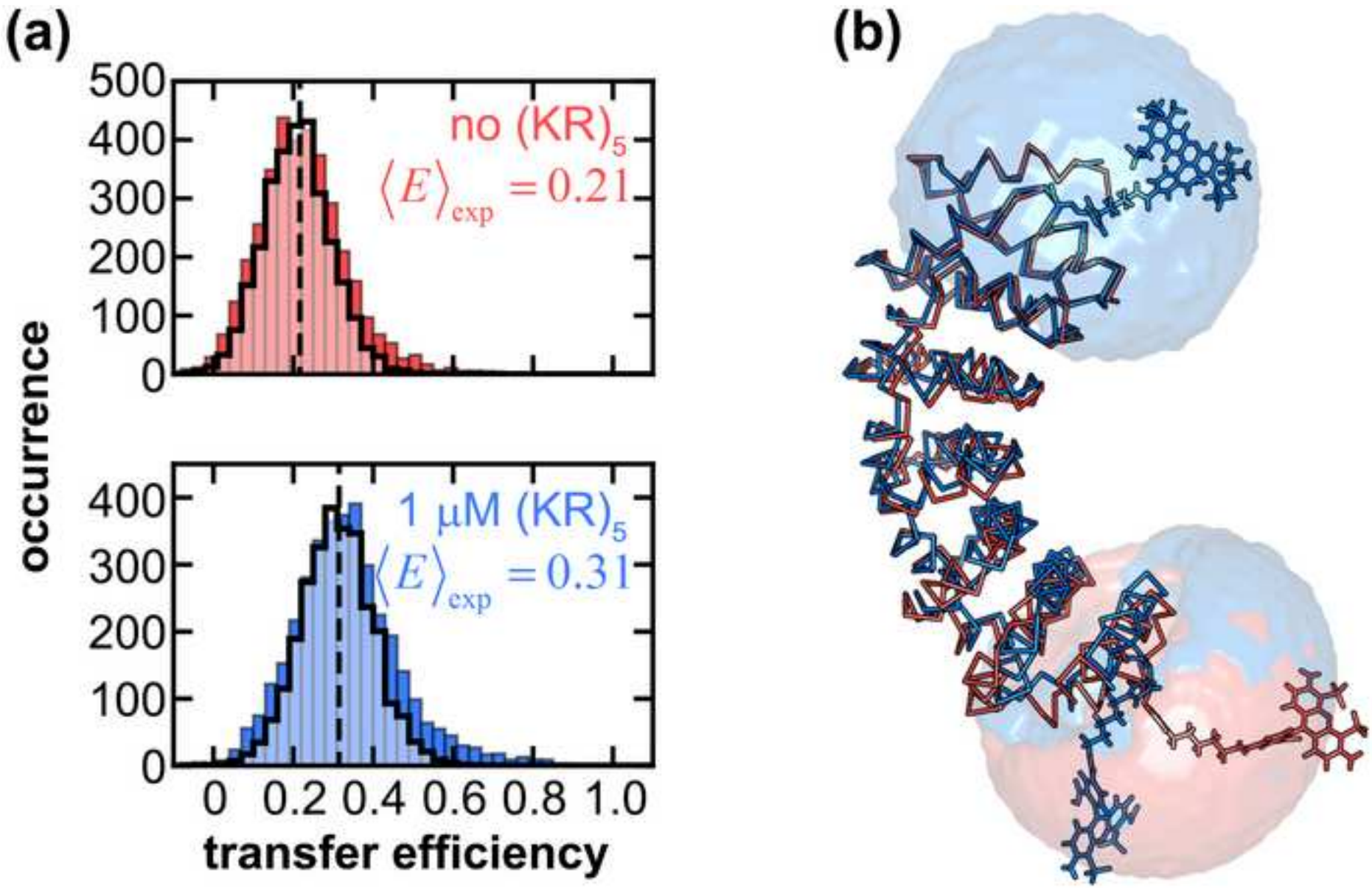

Table 1: Overview of dArmRP structures.

PDB-ID	design type	dArmRP construct	domain-swap	peptide ligand		Ca^{2+}-bridged dimer	space group	dArmRPs per AU	resolution (Å)	reference
				type	canonical binding					
4DBA	consensus	$Y_{II}M_3A_{II}$	yes	-	-	no	P1	4	2.40	Madhurantakam et al., 2012
4DB9	consensus	$Y_{III}M_3A_{III}$	no	-	-	no	$C222_1$	6	2.40	Madhurantakam et al., 2012
4DB6	consensus	$Y_{III}M_3A_{II}$	no	-	-	no	I222	1	1.80	Madhurantakam et al., 2012
4DB8	consensus	$Y_{II}M_4A_{II}$	yes	-	-	no	$P2_1$	4	2.50	Madhurantakam et al., 2012
4V3Q	consensus	$Y_{III}M'_4A_{II}$	no	-	-	yes	$P3_2$	4	2.00	Reichen et al., 2016a
4V3O	consensus	$Y_{III}M''_5A_{II}$	no	-	-	yes	$P4_1$	4	2.00	Reichen et al., 2016a
4PLR	consensus	$Y_{III}M''_5A_{III}$	no	-	-	no	$I4_1$	2	2.10	Reichen et al., 2014b
5MFB	computational	$Y_{III}(Dq)_4CqI$	no	-	-	no	C2	2	2.30	this publication
4PLS	consensus	$Y_{III}M''_5A_{III}$	yes	-	-	yes	$P2_12_12_1$	4	2.35	Reichen et al., 2014b
4PLQ	consensus	$Y_{III}M''_5A_{III}$	no	-	-	no	$P2_12_12$	1	2.10	Reichen et al., 2014b
5AEI	consensus	$Y_{III}M_5A_{II}$	no	$(KR)_5$	yes	yes	$C222_1$	3	2.20	Hansen et al., 2016
4V3R	consensus	$Y_{III}M'_5A_{II}$	no	-	-	no	I4	2	1.95	Reichen et al., 2016a
5MFC	consensus	$Y_{III}M_5A_{II}$	no	$(KR)_4$_GFP	partial	no	$C222_1$	2	2.40	Hansen et al., 2017
5MFD	consensus	$Y_{III}M''_6A_{II}$	no	pD_$(KR)_5$	partial	yes	$P6_3$	8	2.30	Hansen et al., 2017
4D4E	computational	$Y_{III}(Dq)_4C_{PAF}$	no	-	-	no	$P2_12_12_1$	2	2.00	Reichen et al., 2016b
5MFE	consensus	$Y_{III}M_5A_{II}$	no	$(RR)_4$	partial	yes	$P2_12_12_1$	4	1.95	this publication
5MFF	consensus	$Y_{III}M_5A_{II}$	no	$(RR)_5$	no	no	$P2_12_12_1$	4	1.90	this publication
5MFG	consensus	$Y_{III}M_5A_{II}$	yes	$(RR)_4$	no	yes	$P6_2$	4	1.87	this publication
5MFH	consensus	$Y_{III}M_5A_{II}$	no	$(RR)_5$	partial	yes	$P2_12_12_1$	4	2.00	this publication
5MFI	computational	$Y_{III}(Dq.V2)_4CqI$	no	$(KR)_4$	partial	no	$P2_12_12_1$	2	1.45	this publication
5MFJ	computational	$Y_{III}(Dq.V2)_4CqI$	no	$(KR)_5$	partial	no	$P2_12_12_1$	2	1.53	this publication
5MFK	computational	$Y_{III}(Dq.V1)_4C_{PAF}$	no	$(KR)_4$	partial	no	$P2_1$	2	2.30	this publication
4D49	computational	$Y_{III}(Dq.V1)_4C_{PAF}$	no	$(RR)_5$	partial	no	$P2_1$	4	2.10	Reichen et al., 2016b
5MFL	consensus	$(KR)_5$_GS_{10}_$Y_{III}M_6A_{II}$	no	$(KR)_5$_GS10	yes	yes	$P3_221$	3	2.50	this publication
5MFM	consensus	$Y_{III}M_6A_{II}$_GS_{11}_$(KR)_5$	no	GS11_$(KR)_5$	yes	yes	$P2_1$	6	2.30	this publication
5MFN	consensus	$Y_{III}M_5A_{II}$	no	-	-	yes	$P6_322$	2	2.80	this publication
5MFO	consensus	$Y_{III}M_3A_{III}$	no	-	-	yes	$P4_3$	6	1.30	this publication

Structure Validation Reports

Click here to download Supplementary Material (Online Only): Structure_Validation_reports-pdf.pdf

Supplementary Material (Online Only)

Click here to download Supplementary Material (Online Only): SI_raw_data.xlsx

Supplementary Material (to Production)

Click here to download Supplementary Material (to Production): dArmRP_curvature_Supplement_revised_final2.pdf